

Evidence for precategorical extrinsic vowel normalization

Matthias J. Sjerps · James M. McQueen · Holger Mitterer

Published online: 4 January 2013
© Psychonomic Society, Inc. 2012

Abstract Three experiments investigated whether extrinsic vowel normalization takes place largely at a categorical or a precategorical level of processing. Traditional vowel normalization effects in categorization were replicated in Experiment 1: Vowels taken from an [ɪ]–[ɛ] continuum were more often interpreted as /ɪ/ (which has a low first formant, F_1) when the vowels were heard in contexts that had a raised F_1 than when the contexts had a lowered F_1 . This was established with contexts that consisted of only two syllables. These short contexts were necessary for Experiment 2, a discrimination task that encouraged listeners to focus on the perceptual properties of vowels at a precategorical level. Vowel normalization was again found: Ambiguous vowels were more easily discriminated from an endpoint [ɛ] than from an endpoint [ɪ] in a high- F_1 context, whereas the opposite was true in a low- F_1 context. Experiment 3 measured discriminability between pairs of steps along the [ɪ]–[ɛ] continuum. Contextual influences were again found, but without discrimination peaks, contrary to what was predicted from the same participants' categorization behavior. Extrinsic vowel normalization therefore appears to be a process that takes place at least in part at a precategorical processing level.

Keywords Speech perception · Categorization · Psycholinguistics

When listening to speech, listeners are faced with the problem that any particular phoneme is never realized twice in exactly the same way. The production of a speech sound can vary due to factors such as a speaker's gender or accent, or due to the phonetic context in which a phoneme is uttered (Heinz & Stevens, 1961; Hillenbrand, Clark, & Nearey, 2001; Hillenbrand, Getty, Clark, & Wheeler, 1995; Purnell, Idsardi, & Baugh, 1999). The variation that arises as a result of these factors can be so severe that phonemes can overlap with respect to their most important auditory cues (such as the first two formants, F_1 and F_2 , in the case of vowels). Under normal listening conditions, however, listeners are hardly bothered by this variation. Part of the solution to this apparent contradiction lies in the fact that speech is perceived relative to general voice characteristics such as a speaker's pitch (R. L. Miller, 1953; Nearey, 1989) and higher formants (Nearey, 1989). The perception of vowels is influenced not only by vowel-intrinsic aspects (such as pitch and higher formants) but also by vowel-extrinsic context (Ladefoged & Broadbent, 1957). Listeners interpret vowels relative to voice characteristics that are revealed in a preceding sentence. If the speaker has a relatively high F_1 , listeners interpret more ambiguous [ɪ]–[ɛ] sounds as representing /ɪ/ (the vowel with the relatively low F_1), whereas more vowels are interpreted as representing /ɛ/ when the speaker has a generally low F_1 . This study investigates the cognitive locus of listeners' ability to use extrinsic information to compensate for a speaker's vocal tract characteristics.

Johnson, Strand, and D'Imperio (1999) demonstrated that listeners' categorizations of vowels can also be changed through more abstract knowledge such as perceived gender. They showed that categorization behavior for an F_1 [ʊ]–[ʌ] continuum differs depending on whether listeners saw a movie of a male or a female speaker (females generally have higher F_1 values than males). Moreover, they also showed that a similar influence can be found when listeners are made to believe that they are listening to a female or a male speaker

M. J. Sjerps (✉) · H. Mitterer
Max Planck Institute for Psycholinguistics, Wundtlaan 1,
6525 XD Nijmegen, The Netherlands
e-mail: matthias.sjerps@mpi.nl

J. M. McQueen
Behavioural Science Institute and Donders Institute for Brain,
Cognition and Behaviour, Centre for Cognition,
Radboud University Nijmegen, Montessorilaan 3,
6525 HR Nijmegen, The Netherlands

(through instructions). Such results suggest that listeners perceive vowels relative to a cognitive model of the expected vowel space of a particular speaker. Normalization effects would then be the result of a speaker-dependent restructuring of the cognitive vowel space. The idea of such a restructuring of category boundaries contrasts with another proposal about vowel normalization. This proposal suggests that extrinsic vowel normalization has a mainly auditory basis (Watkins, 1991) and takes the place at an auditory level of processing. In this approach, the normalization process is formulated as a mechanism that changes the perception of vowels by taking the overall average spectral shape of a speaker's context sentence and inversely filtering new auditory input for that average (Watkins, 1991; Watkins & Makin, 1994). As a result, the context-dependent adjustment of the auditory representation leads to the context-dependent perception of phoneme identity.

In the present study, we investigated whether vowel normalization effects like those reported by Ladefoged and Broadbent (1957) are best explained by a speaker-dependent restructuring of vowel space or by a change that takes place earlier in the processing hierarchy, at a level of more basic precategorical representations. A speaker-dependent restructuring of vowel space would imply a context-dependent change in the location of phoneme category boundaries. A change in precategorical representations would imply a context-dependent change in the auditory coding preceding phoneme categories. If normalization is the result of a shift in category boundaries, this should not influence the auditory coding that precedes the activation of phonemic categories. We therefore set out to evaluate the two hypotheses by testing extrinsic normalization in a discrimination task that does not rely on the use of phoneme categories (Gerrits & Schouten, 2004), and by comparing discrimination behavior with categorization behavior (Clarke-Davidson, Luce, & Sawusch, 2008; Kingston & Macmillan, 1995; Mitterer, Csépe, & Blomert, 2006).

We examined the locus of extrinsic vowel normalization by comparing listeners' performance on categorization and discrimination tasks, using the same stimuli across tasks. Comparing categorization with discrimination behavior has a long history in research on the perception of speech sounds. Recently, Clarke-Davidson et al. (2008) relied on this method of comparison to argue against a bias interpretation of perceptual learning in speech perception. They showed that the shift in categorization behavior that is traditionally found in perceptual-learning research is accompanied by a shift in the discrimination peaks (using a range of sounds from [s] to [ʃ]). They thus elaborated on the classic findings on categorical perception (Liberman, Harris, Hoffman, & Griffith, 1957) that had been based on the comparison of categorization and discrimination behavior. Liberman et al. showed that boundaries in categorization were accompanied by a peak in

discrimination behavior (i.e., better discrimination at category boundaries) and argued that the discrimination peaks that appeared at phoneme category boundaries reflected the specialized processing of a speech perception module.

This classic finding has attracted a great deal of subsequent research, but this has often shown that the strong form of categorical perception, in which within-category differences are unperceivable, is not tenable (Harnad, 1987). In a similar vein, Gerrits and Schouten (2004) showed that the degree to which speech perception appears to be categorical is related to the task that is used. They observed that discrimination, without discrimination peaks at category boundaries, can be found in some but not all discrimination tasks. Gerrits and Schouten investigated two different discrimination tasks: 2I2AFC (two-interval, two-alternative forced choice), in which on every trial a participant listens to two sounds and has to decide whether the order was AB or BA, and 4I-odddity (or 4I2AFC: four-interval, two-alternative forced choice), in which on every trial a participant listens to four sounds, containing three standards (S) and one deviant (D). The listener has to decide whether the order was SDSS or SSDS by indicating whether the deviant stimulus was in the second or third position. Gerrits and Schouten used the same stimuli in both tasks and found that 2I2AFC gives rise to a discrimination peak at category boundaries, while 4I-odddity does not. They argued that 2I2AFC still induced categorical processing because listeners partly relied on phonetic labels. With 4I-odddity, however, listeners were encouraged to rely on an auditory level of representation.

It has also been found that consonantal stimuli more often give rise to discrimination peaks than do vowels (Fry, Abramson, Eimas, & Liberman, 1962). This is possibly a result of the fact that consonant information is transient, and therefore listeners often have only their category labels left to rely on. For vowels, which are often longer and more stationary, it is easier to focus on auditory information, because vowel cues are stretched over longer time spans (Pisoni, 1973, 1975). Discrimination peaks can nevertheless sometimes be found with vowels (Repp, Healy, & Crowder, 1979). Using a same-different (AX) task, Repp et al. showed that vowel discrimination performance was predicted quite well by vowel categorization behavior, but only if the category labels were obtained in response to the stimulus pairs used in the discrimination task. This suggests that performance on the AX discrimination task, like performance on the 2I2AFC task, may be mediated at least in part by phonetic category labels.

These prior findings suggested that using a 4I-odddity task with vowel stimuli, as opposed to AX or 2I2AFC tasks, would encourage participants to focus on the auditory properties of the sounds. This task therefore seems to be a suitable way to test the level at which extrinsic vowel normalization occurs. This is the approach that we took in Experiment 2. One obstacle remained, however: The 4I-odddity task requires relatively short stimuli, because listeners have to make decisions

about sets of four stimuli. The purpose of Experiment 1 was, hence, to show that extrinsic vowel normalization can be obtained with relatively short stimuli. To achieve this, we tested listeners on stimuli on the continuum from [ɪpapu] to [ɛpapu]. The [papu] part was manipulated to have either a generally high F_1 or a generally low F_1 . Unlike most of the previous experiments on extrinsic vowel normalization, which have used sentence-length contexts (Broadbent & Ladefoged, 1960; Darwin, McKeown, & Kirby, 1989; Ladefoged & Broadbent, 1957; Sjerps, Mitterer, & McQueen, 2011a; Watkins, 1991; Watkins & Makin, 1994, 1996), the present contexts thus consisted of only two syllables. Verbrugge, Strange, Shankweiler, and Edman (1976) did use short contexts, but they found no significant changes in vowel identification. With our Experiment 1, we thus sought to establish the extent to which the stimuli required for Experiment 2 could induce vowel normalization in a categorization task.

Within the 4I-oddity trials that were planned for Experiment 2, a sequence of four trisyllables from the [ɪpapu]-to-[ɛpapu] continuum would be presented. The [papu] part following vowel x was intended to provide the preceding context for vowel $x + 1$. This method was important because it made it possible to use a silent interval (500 ms) between the contexts and the subsequent target vowel, while also providing every vowel with a similar amount of contextual influence. The silent interval was necessary because shorter intervals lead to increasing peripheral auditory influences (Summerfield, Haggard, Foster, & Gray, 1984; Wilson, 1970) that are qualitatively indistinguishable from vowel normalization effects (see also Watkins, 1991). Additionally, this approach made it necessary to present the different speaker conditions in separate blocks, so that the preceding context of the first trisyllable in a quadruplet would have the same preceding context as the second trisyllable. To ensure that the contextual influences of the [papu] part were similar across experiments, this blocked design was also adopted in Experiment 1. The [papu] part in a categorization trial thus provided the context for the first vowel in the subsequent trial and was consistent over the course of a block.

Experiment 1 tested whether these materials and this design could elicit the traditional normalization finding. If these methods indeed resulted in categorization shifts, the same materials could be used in the 4I-oddity discrimination task in Experiment 2.

Experiment 1: Categorization

Method

Participants Ten participants from the Max Planck Institute for Psycholinguistics participant pool were recruited and

tested. They received a monetary reward for their participation. None of the participants reported a hearing disorder, language impairment, or uncorrected visual impairment.

Materials The stimuli consisted of the sequences [ɪpapu] (“ipapu”) and [ɛpapu] (“epapu”). These sequences are meaningless in Dutch. These nonsense words were spoken by a female native speaker of Dutch. The speaker produced the nonwords with the main stress placed on the first vowel. The materials were further processed with Praat software (Boersma & Weenink, 2009). The first vowel of a recording of [ɛpapu] was excised. Using Burg’s linear predictive coding (LPC) procedure in Praat, a filter model was obtained for the vowel by estimating four formants between 0 and 5500 Hz. A source model was estimated with eight prediction coefficients. A range of filter models spanning 200 Hz was created by a linear decrease of F_1 in steps of 10 Hz. These filter models were combined with the source model. The vowels were low-pass filtered between 0 and 1000 Hz and then combined with the higher frequencies of the original vowel (1000–5000 Hz) in order to make the manipulated stimuli sound natural. Filtering was conducted with the standard function in Praat, which operates in the frequency domain (with a smoothing of 100 Hz). All manipulated vowels were adjusted so that their overall amplitudes and their amplitude envelopes matched those of the original vowel. From the original range of possible stimuli, ten target steps were selected, ranging from [ɪ] to [ɛ]. These steps spanned an F_1 range of 180 Hz in steps of 20 Hz (with F_1 values ranging from 190 to 10 Hz lower than the recorded [ɛ]). The original [ɛ] had the following properties: F_0 , 280 Hz; F_1 , 730 Hz; F_2 , 2010 Hz; F_3 , 3494 Hz; voiced duration, ~110 ms (all F values were measured at the midpoint of the vowel). The F_1 of the created vowel continuum thus ranged from ~540 Hz (Step 1 represents [ɪ], which had an F_1 decrease of 190 Hz) to ~720 Hz (Step 10 represents [ɛ], which had an F_1 decrease of 10 Hz). The range of 180 Hz for F_1 that was used here is somewhat larger than the difference between /ɛ/ and /ɪ/ reported for female speakers of Northern Standard Dutch, which is estimated at 136 Hz (Adank, van Hout, & Smits, 2004).

The context part of the stimuli ([papu]) was manipulated by the same procedure, but then with either an increase of F_1 or a decrease of F_1 by 200 Hz for both vowels. The original [papu] context vowels had the following properties: [a], F_0 , 159 Hz; F_1 , 739 Hz; F_2 , 1545 Hz; F_3 , 3631 Hz; voiced duration, ~90 ms; [u], F_0 , 147 Hz; F_1 , 409 Hz; F_2 , 851 Hz; F_3 , 2905 Hz; voiced duration, ~190 ms (all frequencies were measured at the midpoints of the vowel). The voice onset times for the /p/ sounds preceding /a/ and /u/ were 10 and 30 ms, respectively. The manipulated sounds from the [ɪ]-to-[ɛ] continuum were spliced onto the context [papu] parts to create the different steps on the [ɪpapu]-to-[ɛpapu] continuum. This resulted in 20 different stimuli (10 steps \times 2 contexts).

Design and procedure Participants categorized all steps from the [ɪpapu]-to-[ɛpapu] continuum in both F_1 speaker conditions. The F_1 speaker conditions were presented in two separate parts, with the order of presentation of those parts counterbalanced over participants. The ten steps from the continuum were randomly presented in each of 11 blocks within each F_1 speaker part. This resulted in a total of 220 trials per participant (10 targets \times 11 repetition blocks \times 2 F_1 speaker parts), interrupted once by a self-paced pause that separated the two F_1 speaker parts. During the experiment, participants saw the labels “Ipapu” and “Epapu” on a computer screen in front of them. They were asked to identify each stimulus by clicking on the left or the right mouse button. Participants were tested in a soundproof booth, wearing Sennheiser HD 280–13 headphones. The stimuli were presented at a comfortable listening level. The experiment was run using the Presentation software (Version 11.3, Neurobehavioural Systems Inc.).

Results

Figure 1 shows the average categorization results. Responses faster than 100 ms after target onset were excluded (99.3 % of the data were kept). For the repeated measures analyses of variance (ANOVAs), the categorization responses were logit-transformed. The analysis made use of the factors Step (ten steps on the continuum) and Context (low vs. high F_1). The analysis revealed an effect of step, indicating that participants gave more /ɪ/ responses toward the [ɪ] end of the continuum: $F(9, 81) = 75.039$, $p < .001$, $\eta_p^2 = .893$. A comparison between Step 1 (F_1 : 540 Hz) and Step 10 (F_1 : 720 Hz) confirmed that this was due to the fact that more /ɪ/ responses were given toward the [ɪ] end of the continuum: $F(1, 9) = 262.684$, $p < .001$, $\eta_p^2 = .967$. A significant effect was also observed for the factor Context, indicating that participants gave more /ɪ/ responses in the high- F_1 context than in the low- F_1 context: $F(1, 9) = 21.088$, $p < .001$, $\eta_p^2 = .701$. Additionally, an

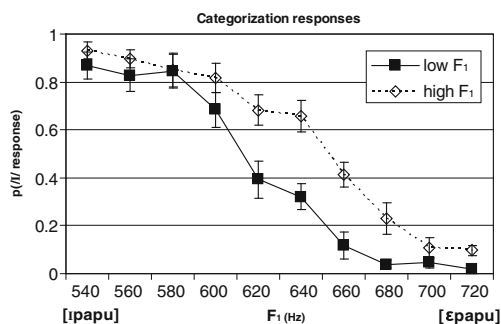


Fig. 1 Experiment 1: Mean probabilities of an /ɪpapu/ response to a range of target sounds from [ɪpapu] to [ɛpapu]. The x-axis displays the target vowel F_1 . The [papu] part was manipulated to have an increased F_1 (high F_1) or decreased F_1 (low F_1) level. Error bars reflect standard errors of the means

interaction was found between the factors Step and Context, reflecting the fact that the effect of context was not equally strong across the continuum: $F(9, 81) = 2.983$, $p = .004$, $\eta_p^2 = .249$. Planned comparisons revealed an effect of context, with more /ɪ/ responses in the high- F_1 context condition, on the following steps: Step 5 (F_1 : 620 Hz), $F(1, 9) = 18.344$, $p = .002$, $\eta_p^2 = .967$; Step 6 (F_1 : 640 Hz), $F(1, 9) = 16.780$, $p = .003$, $\eta_p^2 = .953$; Step 7 (F_1 : 660 Hz), $F(1, 9) = 24.047$, $p = .001$, $\eta_p^2 = .991$; Step 8 (F_1 : 680 Hz), $F(1, 9) = 10.615$, $p = .010$, $\eta_p^2 = .826$; Step 10 (F_1 : 720 Hz), $F(1, 9) = 9.430$, $p = .013$, $\eta_p^2 = .780$.

Discussion

In line with previous findings (Ladefoged & Broadbent, 1957; Watkins, 1991), compensation for a speaker's F_1 characteristics was found with vowel targets on an [ɪ]-to-[ɛ] continuum. The vowels were presented in the context of a speaker with a high or a low F_1 . Listeners categorized more sounds on the [ɪ]-to-[ɛ] continuum as /ɪ/ in the high- F_1 speaker condition than in the low- F_1 speaker condition. This effect was in the expected, compensatory direction. This shows that a relatively short amount of speaker context (i.e., two syllables) can induce shifts in the perception of vowel identity in a compensatory manner in a categorization task. It is important to note, however, that with this design it is uncertain whether the following context also influenced the perception of the target vowels. Moreover, the fact that speaker conditions were presented in a blocked fashion might also have had an influence on the amount of normalization. The exact source of the information leading to the context effect, however, is not important for the present argument. The blocked presentation was necessary for the comparison of these results with those of the following discrimination experiment.

Experiment 2 was set up to investigate whether these compensation effects would also be observed in a task that encouraged listeners to focus on the auditory aspects of the stimuli. The discrimination task that was used was the 4I-oddity task, in which listeners heard sets of three ambiguous standards (S) and one unambiguous deviant (D, either [ɪ] or [ɛ]), in one of two possible orders: SDSS or SSDS. In the high- F_1 speaker condition, an influence of the speaker context should be reflected in lower discriminability for the deviant [ɪ] from the ambiguous standards than for the deviant [ɛ] from the ambiguous standards. This prediction follows from the results from Experiment 1 (see Fig. 1), which show that in the high- F_1 condition, an ambiguous sound that would be used as a standard is perceived as being more similar to [ɪ] than to [ɛ]. In the low- F_1 [papu] context, however, it should be more difficult to discriminate [ɛ] from the ambiguous standards, and easier to discriminate [ɪ]. Because there was a risk that listeners would reach ceiling performance or stay at floor performance

(which could hide any possible context effects), the discrimination experiment was based on both larger (five-step, or 100-Hz) and smaller (four-step, or 80-Hz) step sizes.

A category shift account of normalization (e.g., Johnson et al., 1999) predicts that the normalization effect would arise at a categorical level of processing, and not at a precategorical level. This account predicts that no effect of normalization should be observed in Experiment 2. In contrast, the auditory account (e.g., Watkins, 1991) assumes that the effects of normalization take place at an auditory level, and so should be observed in the 4I-oddity procedure.

Experiment 2: Discrimination

Method

Participants Eight participants from the Max Planck Institute for Psycholinguistics participant pool were recruited and tested. They were selected according to the same criteria that had been used for Experiment 1. None of them had participated in a similar experiment in the year preceding the test, and they received a monetary reward for their participation.

Materials and procedure The stimuli consisted of standard–deviant quadruplets in which the deviant was placed in second or third position, with a 500-ms interstimulus interval (ISI) between the individual trisyllabic sequences. The deviant consisted of one of the endpoints ([I]–Step 1, with an F_1 of 540 Hz, or [ε]–Step 10, with an F_1 of 720 Hz). The standard on each trial was a single step from the middle of the continuum, a vowel with an F_1 of either 620 or 640 Hz (the frequency intervals between two members of a pair were thus either 80 or 100 Hz). These stimuli were created in both the high- F_1 and low- F_1 speaker conditions, resulting in a 2 (deviant: Step 1 or 10) \times 2 (standard: Step 5 or 6) \times 2 (F_1 speaker condition) design.

Stimuli of the resulting eight types were presented in separate subsequent blocks of 32 trials each (i.e., all 32 trials from a particular pair were presented before the next pair was tested). The order of these blocks was randomized across participants. Each of the blocks consisted of four subblocks of eight stimuli. Within a subblock, the numbers of trials on which the deviant would be in second or third position were balanced (i.e., four of each, and the order was randomized).

Participants responded by clicking on either the left button (labeled “2”: i.e., they thought that the deviant was in second position) or the right button (labeled “3”: i.e., they thought that the deviant was in third position) of a button box. Participants received visual feedback (printed “Correct” [correct] or “Fout” [incorrect]) on the computer screen immediately after

each response. This should have provided participants with reinforcement to focus on subphonemic, within-category differences, and as such should have improved their overall discrimination performance.

Note that the participants in Experiment 1 did not receive feedback. We decided not to use feedback in the categorization task in that experiment because feedback could have introduced a bias on subsequent trials, and as such could have influenced the location of the category boundary.

Results

No participant responded earlier than 100 ms after the onset of the second item on any of the trials. Figure 2 shows the average discrimination results. For the repeated measures ANOVA, the discrimination scores were logit-transformed. The analysis made use of the factors Deviant ([I] vs. [ε]), Context (low vs. high F_1), and Standard–Deviant Difference (small [80 Hz] vs. large [100 Hz]). The analysis revealed an effect of standard–deviant difference, indicating that participants were better at discriminating the large standard–deviant steps (circles; see Fig. 2) than the small ones (squares): $F(1, 7) = 25.588$, $p = .001$, $\eta_p^2 = .785$. A marginally significant effect was found for the factor Context, reflecting a tendency for the stimuli in the low- F_1 context (solid lines) to be discriminated better than those in the high- F_1 context (dotted lines): $F(1, 7) = 4.973$, $p = .061$, $\eta_p^2 = .415$. No main effect was found for the factor Deviant: $F(1, 7) = 0.029$, $p = .869$, $\eta_p^2 = .004$. The only significant interaction was that between context and deviant, reflecting the normalization effect that was of main interest: $F(1, 7) = 18.164$, $p = .004$, $\eta_p^2 = .722$. Paired comparisons of the context effect within each deviant showed no effect of context with the [ε] deviant, $F(1, 7) = 0.532$, $p = .489$, $\eta_p^2 = .071$, but there was an effect of context with the [I] deviant, $F(1, 7) = 19.085$, $p = .003$, $\eta_p^2 = .732$. The context effect thus reversed across the two deviants, but, within a vowel, was statistically

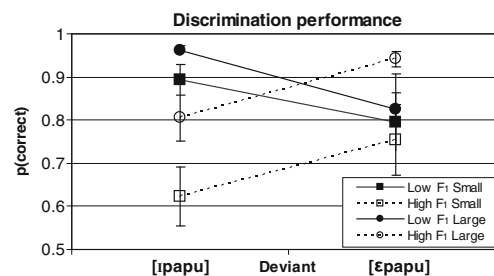


Fig. 2 Experiment 2: Mean probabilities of a correct discrimination response to pairs of stimuli in the 4I-oddity design. Listeners performed a discrimination task in both a low- F_1 (solid lines) and a high- F_1 (dotted lines) speaker condition, defined by the F_1 value in the [pau] part. The deviant stimuli consisted of either [I] or [ε] part, while the standards consisted of the target vowel with an F_1 of either 620 or 640 Hz, making the step size small (squares) or large (circles). Error bars reflect standard errors of the means

significant only for [ɪ]. All other interactions had p values greater than .1.

Discussion

In a 4I-odddity vowel discrimination experiment, we found that listeners' discrimination performance was influenced by the speaker contexts in which vowels were presented. When participants listened to vowels in a low- F_1 speaker context, they found it harder to detect an [ɛpapu] than an [ɪpapu] deviant (with an ambiguous vowel as the standard), whereas in a high- F_1 context, this effect was reversed. This suggests that extrinsic speaker context not only adjusts category boundaries, but also changes precategorical percepts.

One further aspect about the data from Experiment 2 is the surprising tendency for relatively high discrimination scores in the low- F_1 condition with small step sizes (or for that matter, the relatively low discrimination scores in the high- F_1 condition for small step sizes; see Fig. 2). It is unclear what could have led to such an asymmetry. However, in the analysis this asymmetry would have been expressed in the interaction between standard–deviant difference and context, which was nonsignificant. Related to this aspect, the pairwise comparisons showed that the effect of context on discrimination was stronger for the [ɪ]-like deviants than for the [ɛ]-like deviants. One potential reason for the fact that effects are more pronounced at the [ɪ] end of the continua is that the /u/ (the last vowel preceding a target) has an F_1 that is closer to the F_1 of [ɪ] than to that of [ɛ]. The perceptual influence of the F_1 of [u] could therefore be more pronounced on ambiguous tokens close to [ɪ] than on those close to the [ɛ] end of the continuum. This proposal is supported by the observation that the spectral distance between prominences in the context and the target plays an important role in the strength of compensation effects (Laing, Liu, Lotto, & Holt, 2012; Mitterer, 2006).

The finding of the Deviant \times Context interaction supports the idea that vowel normalization takes place at a precategorical level. This is based on the assumption that the 4I-odddity task, especially with vowel stimuli, will lead listeners to focus on the auditory rather than the phonological properties of the stimuli. To validate this assumption, however, it would be necessary to compare actual discrimination patterns with predicted discrimination patterns obtained from the categorization behavior of the same individuals. We therefore tested for extrinsic vowel normalization on the same stimuli, testing the same group of participants on categorization and 4I-odddity discrimination tasks.

Because the categorization task could strengthen the participants' focus on the categorical aspects of the stimuli, the discrimination task was always presented first. In the discrimination task, we tested how well participants were able to discriminate 60-Hz differences across the whole continuum.

That is, the first step (F_1 : 540 Hz) was paired with the fourth step (F_1 : 600 Hz), the second step (F_1 : 560 Hz) was paired with the fifth step (F_1 : 620 Hz), and so on. This provided a discrimination function over the whole range of the continuum, rather than one for only the comparison between the endpoints and midpoints, as had been the case in Experiment 2. In the second part of the experiment, the same participants performed a categorization task with the individual stimuli. This allowed us to establish what the shape of the discrimination functions would look like if participants were to focus on categorical representations.

In this experiment, two questions were at stake. First, does the 4I-odddity task really focus listeners' attention on auditory properties (as the research by Gerrits & Schouten, 2004, had suggested)? Second, does extrinsic vowel normalization influence those auditory properties? The first of these questions could be answered by testing whether the discriminability of stimuli was greater near the category boundary (as established in a categorization task). If such a discrimination peak were absent, it would follow that the discrimination task does indeed focus listeners' attention on auditory properties. The second question could then be answered as follows: If extrinsic vowel normalization influences the auditory properties of stimuli, the low- F_1 context should, in comparison with the high- F_1 context, enhance discrimination performance near the /ɪ/ end of the continuum and inhibit discrimination performance near the /ɛ/ end of the continuum. This follows from the fact that vowels become /ɛ/-like more quickly in the low- than in the high- F_1 context, enlarging the auditory differences at the /ɪ/ end of the continuum for the low- F_1 context relative to the high- F_1 context.

Experiment 3: Predicted versus obtained discrimination

Method

Participants A group of 24 participants from the Max Planck Institute for Psycholinguistics participant pool were recruited and tested. These were selected using the same criteria as in the earlier experiments. None of them had participated in a similar experiment in the year preceding the test. They received a monetary reward for their participation.

Materials and procedure The stimuli that had been created for Experiment 1 were used for the creation of the 4I-odddity pairs. The members of a pair were concatenated to create sets of four stimuli with an ISI of 500 ms. These stimuli consisted of three standard items and one deviant item that was either in second or third position. The combinations consisted of pairs with an F_1 frequency difference of 60 Hz (e.g., Step 1 paired with Step 4, Step 2 paired with Step 5, etc., up to Step 7 paired with Step 10), resulting in seven

pairs. Two versions of these were created, such that both items could occur as either the standard or the deviant (e.g., Step 10 as standard and Step 7 as deviant, or Step 7 as standard and Step 10 as deviant). This resulted in 28 types of discrimination stimuli (7 pairs on the continuum \times 2 choices for the deviant item in a pair \times 2 context speaker F_1 characteristics), with two possible deviant positions (in second or third position). All stimuli were presented eight times, resulting in 224 stimuli in total per participant.

All of the stimuli were presented in seven consecutive blocks of 32 trials each. All trials from a particular pair (e.g., Pair 1, consisting of Step 1 and Step 4) were presented within such a block. The order of the blocks was semi-randomized, such that the average block position of each pair was balanced across participants. After each block, participants were allowed a self-paced pause. Within a block, participants first received all 16 stimuli from one context condition and then the 16 from the other context condition (e.g., for Pair 1, first all stimuli in the low- F_1 context condition and then all stimuli from the high- F_1 context condition). The presentation order of the context conditions was counterbalanced across participants. Within those 16 stimuli, participants first discriminated eight stimuli in one direction (e.g., with Step 1 as the standard and Step 4 the deviant), and then the eight stimuli in the other direction (the orders were balanced across participants). Within those eight stimuli of one type, the position of the deviant (whether it was in second or third position) was balanced and randomized.

The response options were always on the screen (printed as “2” and “3”). Participants were allowed to respond throughout the duration of the trial, which involved pressing one of two buttons on a button box. Participants received visual feedback (printed “Correct” [correct] or “Fout” [incorrect] on the computer screen) immediately after their responses.

For the categorization part of the experiment, the stimuli from Experiment 1 were used (ten steps in two context conditions). Participants categorized subblocks of each of the ten steps (in random order) within a context condition. A subblock with ten steps was then followed by a subblock with stimuli from the other context condition. In this manner, blocks from the two context conditions were alternated until the participants had heard all of the stimuli ten times, resulting in 200 trials. The fact that subblocks with stimuli in the two context conditions alternated across the experiment was therefore the same in both the categorization and discrimination parts.

As in Experiment 1, participants received no feedback during the categorization part. Categorization trials were presented with a fixed ISI of 500 ms. The relatively short ISI meant that the participants had to respond as fast as possible after hearing the vowel (i.e., they were instructed

that they did not have to wait until the [papu] part was over). The discrimination and categorization parts combined lasted around 40 min.

Results

Categorization The top panel of Fig. 3 displays the categorization data. For the categorization task, participants sometimes responded too late because of the restricted time window in which they had to respond (7.5 % of the trials were missed). No responses had to be excluded because they were too fast. For the repeated measures ANOVA, categorization scores were logit-transformed. The analysis made use of the factors Step (ten steps on the continuum) and Context (low vs. high F_1). The analysis revealed a significant main effect of context, indicating that more /i/ responses were given in the high- F_1 condition than in the low- F_1 condition: $F(1, 23) = 13.638, p = .001, \eta_p^2 = .372$. An effect of step reflected the fact of participants' different proportions of /i/ responses toward the [i] versus the [ε] end of the continuum: $F(9, 207) = 112.238, p < .001, \eta_p^2 = .830$. A comparison between Step 1 (F_1 : 540 Hz) and Step 10 (F_1 : 720 Hz) confirmed that this was due to the fact that more /i/ responses were given toward the [i] end of the continuum: $F(1, 23) = 139.404, p < .001, \eta_p^2 = .858$. Finally, an interaction was found between step and context: $F(9, 207) = 2.118, p < .029, \eta_p^2 = .084$. The latter reflects the fact that the strength of the context effect differs across the continuum. Planned comparisons revealed a contrastive effect of context on the following steps: Step 5 (F_1 : 620 Hz): $F(1, 23) = 9.419, p = .005, \eta_p^2 = .291$; Step 6 (F_1 : 640 Hz): $F(1, 23) = 8.206, p = .009, \eta_p^2 = .263$; Step 7 (F_1 : 660 Hz): $F(1, 23) = 11.997, p = .002, \eta_p^2 = .343$; Step 9 (F_1 : 700 Hz): $F(1, 23) = 4.474, p = .045, \eta_p^2 = .163$; Step 10 (F_1 : 720 Hz): $F(1, 23) = 4.524, p = .044, \eta_p^2 = .164$.

Predicted discrimination The middle panel of Fig. 3 displays the average predicted discrimination scores that were obtained from the categorization data. These data points were computed by aggregating over the response data from the categorization part for each of the factors Step, Context, and Participant. Next, the predicted discrimination accuracy was calculated with the equation $\text{Accuracy} = .5 + .5[p(\text{Step A}) - p(\text{Step B})]^2$, where $p(\text{Step A})$ and $p(\text{Step B})$ are the proportions of [ipapu] responses, respectively, for Step A and Step B of a pair (this equation is adapted from Pollack & Pisoni, 1971). This led to a predicted discrimination curve for each participant. For the repeated measures ANOVA, the predicted discrimination scores were logit-transformed. The analysis made use of the factors Step (seven pairs on the continuum) and Context (low vs. high F_1). The analysis revealed a main effect of Step [$F(6, 138) = 11.901, p < .001, \eta_p^2 = .341$], indicating that the predicted discrimination scores

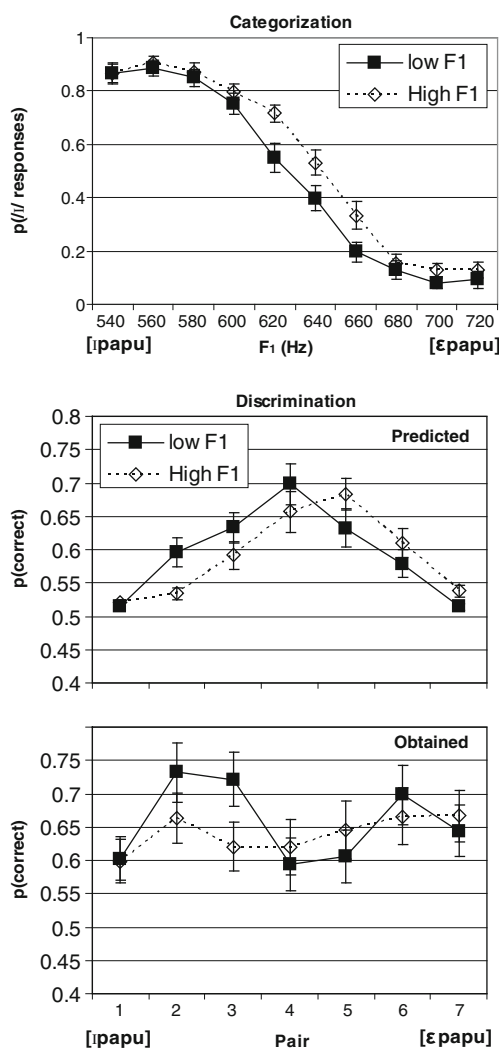


Fig. 3 Experiment 3. (Top panel) Mean probabilities of an /l/ response to a range of target sounds from [ɪpapu] to [ɛpapu]. The [papu] part was manipulated to have an increased F_1 (high F_1) or decreased F_1 (low F_1) level. (Middle panel) Predicted mean probabilities of a correct response to pairs of stimuli in a 4I-odddity design. Discriminability is predicted for “Vpapu” pairs (the V represents the vowel stimulus) for which the F_1 of the initial vowel sounds differ by 60 Hz, sliding across the continuum from [ɪpapu] to [ɛpapu] (e.g., pairs of target vowels with F_1 s of 540 and 600 Hz, 560 and 620 Hz, etc.). This resulted in predicted values for seven steps. Discrimination performance was predicted for both the low- F_1 and the high- F_1 speaker conditions (defined by the F_1 value in the [papu] part). (Bottom panel) Mean probabilities of a correct response to pairs of stimuli in a 4I-odddity design. Scores were obtained for pairs in which the F_1 of the initial vowel sounds differed by 60 Hz, sliding across the continuum from [ɪpapu] to [ɛpapu]. Participants discriminated pairs in both the low- F_1 and high- F_1 speaker conditions. Error bars reflect standard errors of the means

differed across the continuum. Finally, a two-way interaction between Context and Step was found: $F(6, 138) = 2.90$, $p < .011$, $\eta_p^2 = .112$. This reflected the fact that the effect of context was predicted to differ across the continuum. The latter result is a reflection of the contrast effect under investigation.

To investigate the nature of the interaction of pair and context, planned pairwise tests for the factor Context were performed for each pair separately. These revealed an effect of context, with better predicted discrimination in the low- F_1 than in the high- F_1 context for Pair 2 [$F(1, 23) = 11.392$, $p = .003$, $\eta_p^2 = .331$] and a just nonsignificant difference for Pair 3 [$F(1, 23) = 4.258$, $p = .051$, $\eta_p^2 = .156$]. No difference was observed for Pairs 4, 5, and 6. For Pair 6, a just nonsignificant effect of Context emerged, in the opposite direction [Pair 6, $F(1, 23) = 4.117$, $p = .054$, $\eta_p^2 = .152$].

To investigate whether the predicted main effect of step was due to the presence of a significant discrimination peak, separate comparisons of the outermost steps (1 and 7) with the middle step (4) were performed. These revealed that stimuli in the middle of the continuum were predicted to be discriminated better: Step 1 vs. Step 4, $F(1, 23) = 27.403$, $p < .001$, $\eta_p^2 = .544$; Step 7 vs. Step 4, $F(1, 23) = 23.543$, $p < .001$, $\eta_p^2 = .506$.

Obtained discrimination The bottom panel of Fig. 3 displays the actual average discrimination results. No responses had to be excluded because they were too fast. For the repeated measures ANOVA, the discrimination scores were logit-transformed. The analysis made use of the factors Step (seven pairs on the continuum) and Context (low vs. high F_1). The analysis revealed no main effect for the factor Context [$F(1, 23) = 1.910$, $p = .180$, $\eta_p^2 = .077$] and a just nonsignificant effect for Step [$F(6, 138) = 2.078$, $p = .060$, $\eta_p^2 = .083$]. However, a significant interaction between these factors was found [$F(6, 138) = 2.638$, $p = .019$, $\eta_p^2 = .103$]. This reflected the fact that the effect of context differed across the continuum, and as such is a reflection of the contrastive normalization effect.

As in the predicted discrimination data, the nature of the interaction of pair and context was investigated with planned pair-wise tests for the factor Context for each pair separately. These analyses revealed an effect of context, with better discrimination in the low- F_1 than in the high- F_1 context at Pair 2 [$F(1, 23) = 4.521$, $p = .044$, $\eta_p^2 = .164$] and Pair 3 [$F(1, 23) = 8.729$, $p = .007$, $\eta_p^2 = .275$].

Even though the overall effect of step was not significant, we nevertheless tested for the presence of a discrimination peak, with separate tests comparing the overall discrimination scores for the outermost steps (1 and 7) with the score at the middle step (Step 4). These revealed no significant differences: Step 1 versus Step 4, $F(1, 23) = 0.099$, $p = .756$, $\eta_p^2 = .004$; Step 7 versus Step 4, $F(1, 23) = 1.453$, $p = .240$, $\eta_p^2 = .059$.

Predicted versus observed discrimination A final analysis compared the predicted discrimination functions with the obtained discrimination functions. The analysis made use of the factors Step (seven pairs on the continuum), Context (low vs. high F_1), and Test (predicted vs. actual discrimination scores). The analysis revealed that discrimination scores

were better overall in the actual discrimination test than was predicted from the categorization data [$F(1, 23) = 12.820$, $p = .002$, $\eta_p^2 = .358$]. A main effect was also found for the factor Step [$F(6, 138) = 2.809$, $p = .013$, $\eta_p^2 = .109$], and a two-way interaction emerged between the factors Test and Step [$F(6, 138) = 7.197$, $p < .001$, $\eta_p^2 = .238$]. Note that step was significant in the predicted but not in the observed discrimination performance. This interaction thus shows that there was a significant difference between the observed and the predicted discrimination functions, with a peak only for the predicted discrimination performance. That is, the identification functions with the vowel stimuli were steep enough near the category boundary to generate a peak in the predicted discrimination function, but empirically, this peak was not observed.

A two-way interaction was also observed between the factors Context and Step [$F(6, 138) = 4.226$, $p = .001$, $\eta_p^2 = .155$], reflecting the fact that, overall, the discrimination functions for the two context conditions differed. The low- F_1 context, in comparison to the high- F_1 context, enhanced discrimination performance near the /ɪ/ end of the continuum and went in an inhibitory direction near the /ɛ/ end of the continuum.

Discussion

Experiment 3 was set up to investigate two questions. First, does the 4I-odddity task really focus listeners' attention on auditory properties (as the research by Gerrits & Schouten, 2004, had suggested)? Second, does extrinsic vowel normalization influence auditory processes? The first question can be answered affirmatively: The identification functions, even with vowel stimuli, were steep enough to predict a peak in discrimination performance at the category boundary. Such a peak, however, was not observed. Discrimination performance was overall not influenced by step, and differed significantly from the predicted discrimination function.

This allowed us to answer the second question. Again, the answer is affirmative, because the F_1 contexts enhanced and inhibited performance largely as predicted. As in the predicted discrimination function, we found an interaction between the factors Step and Context, such that toward the [ɪ] end of the continuum, listeners were better able to discriminate sounds in the low- F_1 condition, while the effect was in the other direction toward the [ɛ] end of the continuum. Note that these results are very similar to those of Experiment 2. In both experiments, we obtained a Context \times Deviant interaction, and in both experiments, the effects of context on discrimination were somewhat stronger for the [ɪ]-like deviants than for the [ɛ]-like deviants. The latter observation supports our earlier suggestion that the influence of the F_1 of /u/ (which is closer to the F_1 of [ɪ] than to

that of [ɛ]) is more pronounced for ambiguous stimuli that are close to the [ɪ] end of the continuum than for those close to the [ɛ] end of the continuum.

The data presented above showed a reliable dissociation between the effects obtained in discrimination and those predicted from categorization. In principle, this dissociation suggests that in our discrimination task, our listeners relied only on auditory aspects of the stimuli, and not on their categorical aspects. It should, however, be noted that this conclusion would be too strong. To be able to draw such a conclusion, we would also have to show that the context effects that were elicited in the two tasks were of exactly equal size. It is, however, unlikely that the effect sizes were completely equal across the discrimination and categorization tasks. This is related to the ISIs in Experiment 3. The discrimination part of Experiment 3 had fixed, short ISIs (of 500 ms). The stimuli in the categorization part also had fixed, short ISIs of 500 ms. It is possible that short ISIs in the categorization part may have led to a decrease in the strength of the context effects.¹ J. L. Miller and Dexter (1988) investigated the regressive influence of speech rate on the location of the category boundary for the English /b/–/p/ contrast. They presented sounds on a /b/–/p/ continuum in word-initial position and manipulated the speech rate in the following part of the word. They found that when listeners responded quickly, the influence of the speech rate of the rest of the word was smaller than when listeners responded more slowly. In the present experiment, regressive context effects might also have played a role. Since we urged our participants to respond quickly in the categorization part, the categorization data might underestimate the full effect of context (i.e., the context effect detected in the discrimination task). Arguably, with our materials the effect observed by J. L. Miller and Dexter might have surfaced with longer delays. It should be noted, however, that, contrary to the experiment of J. L. Miller and Dexter, our stimuli were presented in a blocked fashion, such that regressive effects would be in the same direction as the progressive effects resulting from the preceding stimuli. It is therefore likely that any additional influence of the regressive effect within a stimulus block would only be minor. However, this particular aspect shows that a direct comparison of discrimination and categorization functions is not straightforward, since performance in each task is under the influence of different factors. This makes it unlikely that effect sizes of context could be fully matched across the experiments. It is important, however, that we observed a qualitative difference between the actual discrimination functions (with no peak at the category boundary) and the functions predicted by the categorization behavior (with a peak at the category boundaries). This shows that, in the

¹ We thank Jim Sawusch for bringing this possibility to our attention.

discrimination part, participants largely focused on the auditory properties of the stimuli, and yet there was still an effect of context. At the auditory level, therefore, we observed effects of extrinsic vowel normalization.

General discussion

In the series of experiments reported here, we investigated the cognitive locus of extrinsic vowel normalization. In the first experiment, we showed that listeners interpret an [I]-to-[ε] continuum differently depending on the F_1 range of a speaker in a context stimulus. This is a replication of earlier findings that have shown that perceived vowel identity can be influenced by a preceding sentential context (Ladefoged & Broadbent, 1957; Sjerps et al., 2011a; van Bergem, Pols, & Koopmans-van Beinum, 1988; Watkins, 1991; Watkins & Makin, 1994, 1996). In the experiments reported here, however, an effect was established using a speaker context that was only two syllables long (/papu/). This context was spliced after the target vowel (targets were of the general structure /Vpapu/). Moreover, there was always a 500-ms interval between the context part of a preceding trial and a target vowel. This approach prohibited potential peripheral auditory influences (Summerfield et al., 1984; Wilson, 1970). Peripheral auditory influences show an effect in the same direction as the normalization effects that were under investigation, and as such are indistinguishable from those effects. However, peripheral auditory effects strongly diminish after a longer ISI. This allowed us to focus on central compensation processes (Watkins, 1991).

In the second experiment, these stimuli were used in a 4I-oddy discrimination experiment. The phoneme category shift proposal (e.g., Johnson et al., 1999) assumes that vowel normalization is the result of a shift of category boundaries, taking place at a categorical level of processing. As such, this proposal predicted that Experiment 2 would not lead to reliable vowel normalization effects. The auditory proposal (Watkins, 1991; Watkins & Makin, 1994, 1996), however, assumes that compensation effects arise at precategorical levels. This proposal thus predicts that vowel normalization effects will express themselves not only in categorization, but also in an auditory-focused task such as the 4I-oddy task. The results confirmed the predictions made by the auditory proposal: Discrimination performance was dependent on speaker context. In the context of a speaker with a high F_1 , listeners found it more difficult to discriminate between an [I] and an ambiguous sound than between an [ε] and an ambiguous sound. In the context of a low- F_1 speaker, this pattern was reversed.

The third experiment confirmed that normalization effects in 4I-oddy discrimination experiments, as used here, are for an important part precategorical in nature. If the present

experiments had failed to encourage listeners to focus on auditory representations (so that they used category labels instead), a discrimination peak should have been found half-way along the continuum. This would have made it impossible to attribute the shifts in discriminability to a precategorical processing stage. Simulated discrimination functions, based on categorization data, showed how a categorical strategy would have been expressed in the discrimination scores. The actual discriminability of vowel pairs across the [I]-to-[ε] continuum was influenced by the speaker context. Unlike the predicted discrimination functions, however, the actual discrimination functions in Experiment 3 did not show a discrimination peak at the category boundary.² Listeners therefore had indeed focused on the precategorical properties of the vowel stimuli.

The simplest account of the normalization effects observed with the same stimuli across all three of the present experiments, therefore, is that the effects reflect, for an important part, a precategorical process. Note, however, that this investigation does not determine whether these precategorical representations already contain some form of abstraction (e.g., abstractions over auditory cues). Nevertheless, this study does show that normalization can take place at a level that precedes the level at which phonemic judgments are made. The fact that a sound that is ambiguous between [ε] and [I] is more often labeled as /I/ in a high- F_1 context and more often as /ε/ in a low- F_1 context (see Fig. 1) thus appears to involve a shift in perceptual space that takes place at a precategorical level. This provides support for the notion that an important component of extrinsic normalization is the result of a process that takes place at an auditory level of representation (Sjerps, Mitterer, & McQueen, 2011a, 2011b; Watkins, 1991; Watkins & Makin, 1994, 1996).

The present findings also provide support for recent findings by Sjerps, Mitterer, and McQueen (2012) and Sjerps et al. (2011b). Sjerps et al. (2012) reported effects of speech and nonspeech contexts on the discriminability of vowels, suggesting that very similar contrastive processes operate across speech and nonspeech materials. Furthermore, using an event-related potential design, Sjerps et al. (2011b) observed very

² Repp et al. (1979) suggested that the correspondence between identification and discrimination performance for vowel stimuli needs to be assessed in the same format, rather than performing an identification task with one stimulus per trial and a discrimination task with two stimuli in a row (or four stimuli in our case). They based their suggestion on the observation of contrast effects when stimuli are identified in the format of an AX discrimination task. They suggested that the identification of the two vowels would be strongly influenced by their adjacent co-presence in a classical AX task. This was unlikely to be an issue in the present case, because the contexts preceding the to-be-identified or to-be-discriminated vowels were identical in both tasks (namely, the [papu] part of the preceding stimulus). Repp et al. also showed that such intervening materials abolish any contrast effect on the target vowels (see their Table 4).

early influences of the auditory properties of contexts on vowel perception (i.e., during the N1 time window, which is associated with precategorical processing). However, in both of those studies, a discrimination task was used, motivated by the presumed (but unverified) auditory nature of that task. The present results validate this assumption, and hence strengthen those findings.

A purely auditory approach, however, is not able to explain the findings reported by Johnson et al. (1999) of nonauditory contextual influences on the categorization of vowel continua. In that study, vowel categorization shifts were induced by asking participants to imagine listening to a male or a female speaker. Learned covariations between speaker gender and average F_1 were shown to influence subsequent categorization. The combination of the findings of Johnson et al. and those reported here suggests that vowel normalization effects do not arise from a single processing level. Normalization effects could, thus, have both precategorical and higher-level cognitive components. The effects observed in Experiments 2 and 3—due to the nature of the 4I-oddity task, as revealed especially by the absence of a discrimination peak at the category boundary in Experiment 3—can mainly be attributed to precategorical processes. But the effect observed in Experiment 1 might have a cognitive component. The important point, however, is that a major component of the effect in a categorization task such as Experiment 1 is likely to be the result of a restructuring of perceptual space at a precategorical level of processing. If this low-level process is at work in the 4I-oddity task, there is every reason to suppose that it is also at work in a categorization task.

Similar approaches have been taken in the closely related fields of compensation for coarticulation (Stephens & Holt, 2003) and compensation for assimilation (Mitterer et al., 2006), with similar results. Mitterer et al. reported on an optional Hungarian assimilation rule: The final [l] in the syllable [bɒl] is produced as [l] in “balnal”:[bɒlna:l] but can be produced as [r] in “balrol”:[bɒrro:l]. Hungarian listeners compensate for this assimilation rule because they perceive more sounds on an [l]-to-[r] continuum as [l] in the assimilation-viable word “balrol” (as compared to the word “balnal”). This does not depend on language background, because Mitterer et al. found that Dutch listeners showed the same compensation effect (Dutch does not have this rule, nor were the listeners familiar with Hungarian). Interestingly, the same effect was found with discrimination: Listeners, both Dutch and Hungarian, found it harder to distinguish between [bɒl] and [bɒr] in the assimilation-viable “. . . rol” context than in the assimilation-unviable “. . . nal” context. Additional findings showed that nonspeech contexts (nonspeech versions of the [ro:l] and [na:l] syllables) induced similar effects. Mitterer et al. argued that their findings support the idea that these kinds of compensation effects take place at an auditory-processing level.

In a similar vein, Stephens and Holt (2003) tested whether compensation for coarticulation generalizes to the influence of speech contexts on nonspeech targets. They elaborated on the finding by Mann (1980) that an ambiguous CV syllable half-way between [ga] and [da] is more often identified as /ga/ (which has a low F_3) when they are preceded by [a] (high F_3), but more often are identified as /da/ (high F_3) when they are preceded by [ar] (low F_3). Stephens and Holt found that the perception of both speech and nonspeech versions of the [ga]–[da] continuum could be influenced by the preceding [a]–versus–[ar] speech context. Because nonspeech targets cannot easily be categorized by participants, Stephens and Holt used a discrimination design for which category labels were not necessary, showing that it was more difficult to discriminate steps on the continuum when the preceding syllables influenced the targets, such that the F_3 values were perceptually brought closer to each other than when the context acted to increase the perceived difference in F_3 . This effect was found for both speech and nonspeech targets. These findings suggest that this type of compensation for coarticulation also has an auditory basis. These reports, which all investigated the contrastive nature of speech perception, therefore all came to the same conclusion: Compensation/normalization processes can largely be attributed to auditory processes because the effects of these processes are visible in discrimination tasks.

A final aspect of the results that deserves consideration is that the discrimination results from Experiments 2 and 3 suggest that the effect of context on extrinsic vowel normalization is not the result of a shift of the complete perceptual space. Rather, the restructuring of perceptual space seems local. That is, the influence of the F_1 manipulation is restricted to frequency regions that are close to the average F_1 in the context. A shift of perceptual space over the complete frequency range should have led to no differences in discriminability over the two context conditions. For example, consider Experiment 3: If vowel normalization were to result in an auditory transformation that shifted all frequencies up by 100 Hz, then the differences between the F_1 s of the stimulus pairs should remain at 60 Hz. Discriminability across the continua should then not have differed for the two context conditions. The finding that normalization reflects local changes dovetails well with data showing that vowel normalization affected only front vowels if the context contained only front vowels, but both front and back vowels if the context contained both types of vowel (Mitterer, 2006).

To conclude, this article reports on extrinsic normalization effects like those first reported by Ladefoged and Broadbent (1957). These effects were established with a paradigm that does not rely on the use of phoneme categories and that reduces possible higher-level influences. Such influences can hide or exaggerate normalization effects by introducing biases that cause shifts in the categorization functions. Here, the comparison between categorization and discrimination behavior

established that extrinsic properties of a speaker's vocal tract influence vowel perception at an early, precategorical level of processing.

Author note We thank Susanne Brouwer for recording the stimuli, along with Jim Sawusch and an anonymous reviewer for constructive feedback

References

- Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America*, 116, 1729–1738.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer [Computer program]. Retrieved from www.praat.org
- Broadbent, D. E., & Ladefoged, P. (1960). Vowel judgements and adaptation level. *Proceedings of the Royal Society B*, 151, 384–399.
- Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, 70, 604–618.
- Darwin, C. J., McKeown, J. D., & Kirby, D. (1989). Perceptual compensation for transmission channel and speaker effects on vowel quality. *Speech Communication*, 8, 221–234.
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171–189.
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66, 363–376. doi:10.3758/BF03194885
- Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1–52). New York, NY: Cambridge University Press.
- Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 33, 589–596. doi:10.1121/1.1908734
- Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109, 748–763. doi:10.1121/1.1337959
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111.
- Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27, 359–384.
- Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, 97, 1261–1285. doi:10.1121/1.412169
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.
- Laing, E., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology*, 3, 203.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407–412.
- Miller, R. L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America*, 25, 114–121.
- Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 369–378. doi:10.1037/0096-1523.14.3.369
- Mitterer, H. (2006). Is vowel normalization independent of lexical processing? *Phonetica*, 63, 209–229.
- Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology*, 59, 1395–1424. doi:10.1080/17470210500198726
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85, 2088–2113.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253–260. doi:10.3758/BF03214136
- Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, 3, 7–18. doi:10.3758/BF03198202
- Pollack, I., & Pisoni, D. (1971). Comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 24, 299–300.
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18, 10.
- Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 129–145.
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011a). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics*, 73, 1195–1215.
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011b). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*, 49, 3831–3846.
- Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2012). Hemispheric differences in the effects of context on vowel perception. *Brain and Language*, 120, 401–405.
- Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech. *Journal of the Acoustical Society of America*, 114, 3036–3039.
- Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics*, 35, 203–213.
- van Bergem, D. R., Pols, L. C. W., & Koopmans-van Beinum, F. J. (1988). Perceptual normalization of the vowels of a man and a child in various contexts. *Speech Communication*, 7, 1–20. doi:10.1016/0167-6393(88)90018-0
- Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talkers vowel space. *Journal of the Acoustical Society of America*, 60, 198–212.
- Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, 90, 2942–2955.
- Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *Journal of the Acoustical Society of America*, 96, 1263–1282.
- Watkins, A. J., & Makin, S. J. (1996). Effects of spectral contrast on perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, 99, 3749–3757.
- Wilson, J. P. (1970). An auditory after-image. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 303–318). Leiden, The Netherlands: Sijthoff.